A NEW GLOBAL OPTIMIZATION METHOD FOR ELECTRONIC CIRCUIT DESIGN

by

A. Groch, L,M. Vidigil, £ 3.tf. Director

Decamber, 1932

DRC-13-49-32

A NEW GLOBAL OPTIMIZATION METHOD FOR ELECTRONIC CIRCUIT DESIGN*

by

A, Groch--

L. M. Vidigal***

S. W. Director

Department of Electrical Engineering
Carnegie-Mellon University
Pittsburgh, PA 15213

ABSTRACT

This paper describes an algorithm which can determine a neighborhood of the global optimum of an objective function as well as an estimate of the global optimum. Given this information, a local optimization procedure can be employed to locate the global optimum. The utility of this algorithm is demonstrated by several examples.

---

*This work has been supported in part by the National Science Foundation under Grant 78-25755 and ITT, Inc.

--On leave from the Technical University of Gdansk, 6dansk, Poland

***Now at the Technical University of Lisbon, Lisbon, Portugal

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Numerous papers have been published which describe the use of optimization methods to improve the performance of electronic circuits (see for example [1]$_9$ [2], [3], and [4]). One of the often-heard complaints about these approaches is that the resulting "optimum solution*[9] is in fact only a local optimum. In this paper we develop a method which can be used to find the neighborhood of the global optimum. Once this neighborhood has been found a local optimization procedure can be employed, with a starting point in this neighborhood, to find the global optimum.

While some work has been reported in the area of global optimization, most methods require lengthy computations and a large number of objective function evaluations [S3. (Note that each objective function evaluation implies a computationally expensive circuit simulation). For these reasons, these methods have not been used for electronic circuit design. However, satisfactory results may be obtained from heuristic approaches which employ a relatively low number of objective function evaluations [51. Typical of such methods are the probabilistic methods introduced by Kushner [63, [73 and Zilinskas [83. Although Kushner's and Zilinskas* methods require only a small number of objective function evaluations, they were developed to solve one-dimensional problems only, clearly not of much use for optmizing circuits in which several designable parameters are usually required.

In this paper we present a new technique for global optimization which, while based on the concepts introduced by Kushner and Zilinskas. is simpler, faster and extendable to higher dimensions. We begin our discussion in Section 2 by formulating a general one-dimensional global optimization problem, and then describe the Kushner and Zilinskas methods for its solution. In Section 3 we introduce, the new simplified algorithm and extend it to higher dimensions in Section 4. In Section 5 several numerical examples illustrate the behavior of this algorithm. Finally, in Section 6 we discuss some areas for future development

2. ONE-DIMENSIONAL PROBABILISTIC METHODS

Let f(x), x € ACR", f:R°->R^{1} be a real continuous function, where A denotes the feasible region which is assumed to be closed and simply connected. Although f may have several local minima, we assume that there exists an x^{0}* such that

$$Hx"*) \; Z \; f(x), \qquad V x f A \tag{1}$$

We define an *e*-neighborhood of x^{**1}, M_{ϵ} as [93

$$M_\epsilon = \{x | x \ll A \text{ and } \|x - x^{\wedge}\|^{\wedge} f\} \quad (2)$$

The object of a global optimization procedure is to find a point x* in M_ϵ which, when used as the initial starting point for a local optimization,* results in the determination of x^{opl}. In what follows we assume that A is a hyperrectangular whose vertices are defined by the so-called box constraints on the designable parameters

$$x_j^L \leq x_j \leq x_j^U \quad (3)$$

where x_j is the jth component of x.

At this point we limit our considerations to the one dimensional case, i.e., we assume that x is a scalar. Now consider the global optimization methods proposed by Kushner [7] and Zilinskas [8]. In both these methods it is assumed that the real, continuous objective function f(x), frR'-^R^1, is a particular realization of some stochastic process f(x,<*), with x € A = [a,b]CLR* and *w* < Q, where Q is a set of random events. This assumption implies the a priori selection of a stochastic process f(x,i*) and the definition of a measure for a search procedure in a feasible region. It has been shown by Kushner [7], Zilinskas [8], Mockus [10], and later by Archetti [11], that it is convenient, and well justified, to assume f(x) to be a sample path of a Wiener process in [a,b]. Several factors leading to this conclusion [7], [10] are: that it is intuitively related to the physical functions: that it is a process of Gaussian, independent and infinitely divisible increments; and that it satisfies conditions of continuity of its realizations and homogeneity in a given region of A. Furthermore, as will be seen below, if f(x,») is a Wiener process, the behavior of the objective function, f(x,<*), conditioned by previous function evalautions performed, can be characterized in probabilistic terms by exceptionally simple formulas.

The major difference between the two methods is the sampling policy which is used in the feasible region to determine the next point at which a function evaluation is to be performed. Let $x^{0\wedge 1}$ denote the estimate of the solution x'^{*11} in the region [a,b] after k steps of the optimization procedure, and $fix^{0*1*})$ the objective function value at this point. In the Kushner method the next point at which a function evaluation is to be performed is the point that

*In general, local optimízatíon methods are more efficieni ín the neighborhood of the optimum [5].

maximizes.

$$P\{[f(x,\omega) \leq [f(x^{opt,k}) - \delta_k]\} \tag{4}$$

in $[a_v b]_v$ where P denotes probability, and i_k is a series of positive constants. We denote this (possibly nonoptimum) point by x^{**1}. In other words, this procedure chooses the point that is most likely to be the competitor with $x^{01*'*}$ (for a given $\&J$ for the location of $x^{''1''}$. Although the resulting formulas which are used to determine X^{**1} are relatively simple, the Kushner method is not very useful for practical applications because of the rather arbitrary choice of the positive constants $\&_k$. The suggestions given by Kushner in this regard are insufficient to aid the designer in this quest for x^{opt}.

In the Zilinskas method, which belongs to a group of methods known as Bayesian methods [12] [131, the next point at which a function evaluation is made is taken to be the point which is the solution of the problem

$$\max_{x \in A} f_{k*1}(x) \; ^{s} \; \max_{XfA} ttW\wedge)^{t} — B\{\min(f(x^{opl*\wedge_v v})_v \; f(x,«*))|zM] \tag{5}$$

where £ denotes the expected value, and z* is a vector which contains the information gained about process after first evaluations of f(x).

$$z^* = [fOc^1).\ f(r^1). \ldots .f(x^k).\ x^!, \ldots \wedge]^T \tag{6}$$

The basic philosophy behind this approach is that at every stage an average improvement towards the minimum of f(x) is maximized, It follows from (5) that $\wedge,(x)$ corresponds, for a given x, to

$$\int_{-oo}^{f(x^{opt,k})} F(u)\, du \; - \tag{7}$$

where F(u) is the distribution function of the random variable u = min ($f(x^{\circ ruk})$, f(x,«)),

$$F(u) = \begin{cases} F(v) & \text{for } v < f(x^{opt,k}) \\ 1 & \text{for } v \geq f(x^{opt,k}) \end{cases} \quad (8)$$

where random variable v corresponds to f(x.») for a given x.

Since a Wiener process was assumed, we obtain from (5), (6). and (7) the expression

$$ttJx\wedge Jx) \int_{-00}^{\zeta} \left[\frac{}{**} \int_{-00} e^{-2}\text{«dt]} \, du \quad (9)$$

where

$$C = \frac{f(x^{opt,k}) - m_k(x)}{C<r_k(x>}$$

$m_k(x) = E\{f(x,\omega) \mid z^k\}$

$\sigma_k^2(x) = Var\{f(x,\omega) \mid z^k\}$

$m_1(x) = f(a)$

$a]ix) = \sigma^2 x$

$f(x^{opt,1}) = f(a)$

It is convenient at this stage in the procedure to order the observation points (Le. those points at which the objective function has been evaluated) into a monotonically increasing sequence

$$a = x^{1,k} \leq x^{2,k} \leq \ldots \leq x^{k,k} = b \quad (10)$$

1 where each $x^{>k}$ is one of the observation points xV i=l,..,k in (6). Observe that as new observation points are determined, complete reorderings may be necessary, Le., $x\wedge^k$ may be

different from x^{*kM}.

Using the well known properties of a Wiener process [10] we may easily obtain the conditional expected value and the conditional variance in each subinterval $[x^{i-1,k}, x^{i,k}]$, (i=2,3,...,k) of the region A = [a,b] from the following equations:

$$m_k(x) = \frac{f(x^{i-1,k})\,(x - x^{MJC}) \cdot f(x^{M,k})\,(x^{i,k} - x)}{x^{u} - x^{wjL}}$$

and

$$\sigma_k^2(x) = \sigma^2 \frac{(x - x^{1u1\wedge})\,(x^{i,k} - x)}{x^{i,k} - x^{i-1,k}} \qquad (11)$$

where $x * [x^{1,,uk}, x^{i,k}]$. From (11) we observe that $m_k(x)$ is a piecewise linear function and the equation for $<r^{\wedge}(x)$ describes a $^{\wedge}$ uadratic on every subintervaL Observe that outside of the region A, $m_k(x)$ becomes the constant values f(a) and f(b) and the variance $<\hat{r}_k^2(x)$ becomes $\epsilon r^2 1\, x\text{-}a\,|$ and $<r^2|x\text{-}b|$, respectively, (see Fig. 2-1). a^2 is a parameter characteristic of the Wiener process. Its value must be estimated for each problem. For this purpose an unbiased maximum likelihood estimator is used. [103. Estimation of a^1 requires M initial, arbitrary chosen observations (cq. M = 6).

To determine the next observation point $x^{k*\backslash}$ the maximum values of $^{\wedge}_M(x)$ in each subinterval have to be found and then compared. This task is made easier by the following properties of $^{\wedge}_1(x)$:

1. it is unimodal in each subinterval

1 it is an increasing function of a and nondecreasing function of the subinterval length

3. it is a nonincreasing function of the difference $m_k(x) - f(x^{optvk})$

The (k+l)st estimate of the optimum x^{0*}, denoted as x^{opuk+1} is the point at which the conditional expectation ECftolz**[1]} is minimum. Since this is a piecewise linear function of x, $j?**+^{l}$ is the point at which $f(i^{ijL*!})$ (i=l,2,...,k+l) is a minimum. Although the proof of convergence to the global optimum [8Has in the Kushner method) requires in the limit that k->oo, usually no more than about 20 observations are needed to obtain a point which lies in the neighborhood of the minimum (at least locally.** Since the termination criterion proposed

C

Doc to the inadequacy of the stochastic model for the objective function fU). when distances between subsequent observation* arc small, deierministics local optimization methods are more efficient in the neighborhood of the minimum.

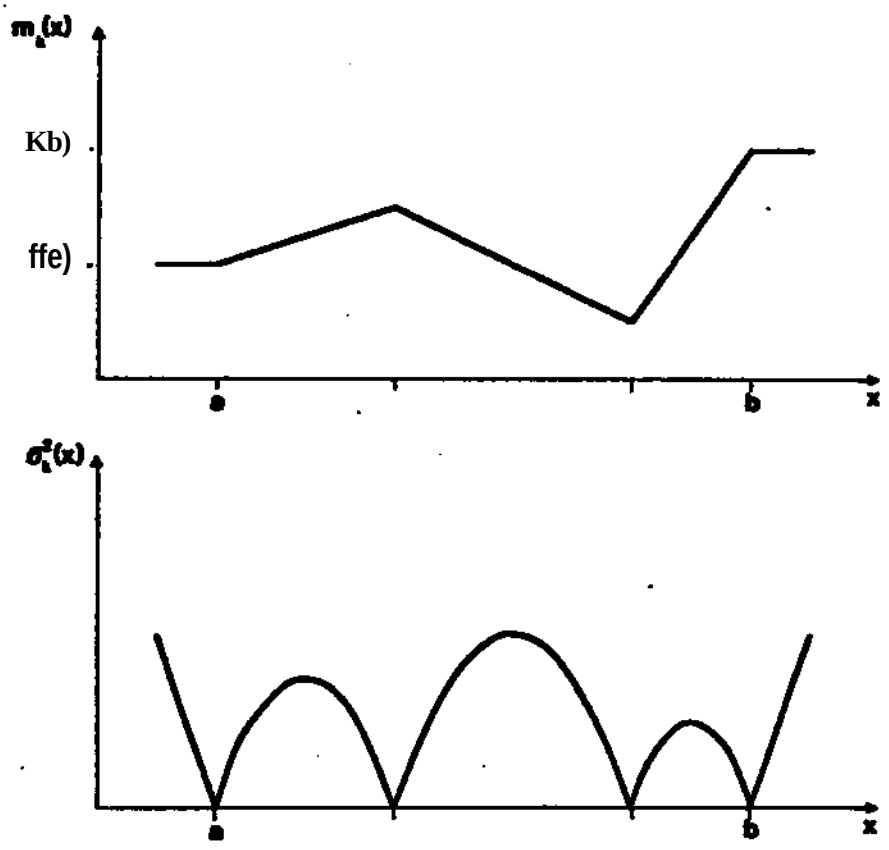

Figure 2-1: Expected value and variance of a Wiener process: a conditional and an unconditional

by Zilinskas [81 is not reliable, usually the above procedure is carried out for a predetermined number of iterations.

The efficiency of the Zilinskas algorithm depends on a constant c £ 1, which is associated with the variance of the Wiener process (9). For large values of c the method becomes more "global" but usually more iterations are needed to locate the neighborhood of the minimum. Note that for small values of c we increase the risk that the final solution may correspond to a local minimum only. A typical shape for $\wedge_M(x)$ is shown in Fig. 2-1

Solution of one-dimensional global optimization problems, by either the Kushner or Zilinskas methods usually requires only a small number of objective function evaluations [7], [81, [10]. The reason for this is as follows. In practice, a multimodal objective function f(x) is continuous and smooth. Hence, there is some correlation between function values at points which are close to one another. Thus, it is reasonable to choose as the new observation point, X^{**1}, either a point close to the current best estimate of the global optimum $x\wedge*$ or in the interval with the largest uncertainty (the longest one). Both the Kushner and Zilinskas methods make certain tradeoffs between these two goals [7], [8], [14]. Different computational steps, resulting from the different sampling policies employed, are determined by the same conditional parameters of a Wiener process (11) which in turn depends on the length of subintervals and values of f(x) at their ends (see $\wedge_k+,(x)$ in Fig. 2-2). Unfortunately as indicated above, both methods have limited utility when applied to circuit design problems. The Kushner method requires arbitrary choices of certain constant values (see Section 2) and the Zilinskas method has, at every stage in the procedure, a significant computational overhead related to the solution of (9).

3. A NEW APPROACH TO ONE-DIMENSIONAL GLOBAL OPTIMIZATION

We now develop a new probabilistic approach to the one-dimensional global optimization problem which is faster than Zilinskas' method and more reliable than Kushner's method. Furthermore, this new approach may be easily extended to higher dimensions. While this new method retains the Wiener process as a stochastic model of the objective function, it employs a new sampling policy.

Observe that for any fixed point $\bar{x}$ « A, the distribution of a random variable *f(%w)* is Gaussian with an expected value m(9 and a variance <r(S). With this in mind, we introduce a new auxiliary function

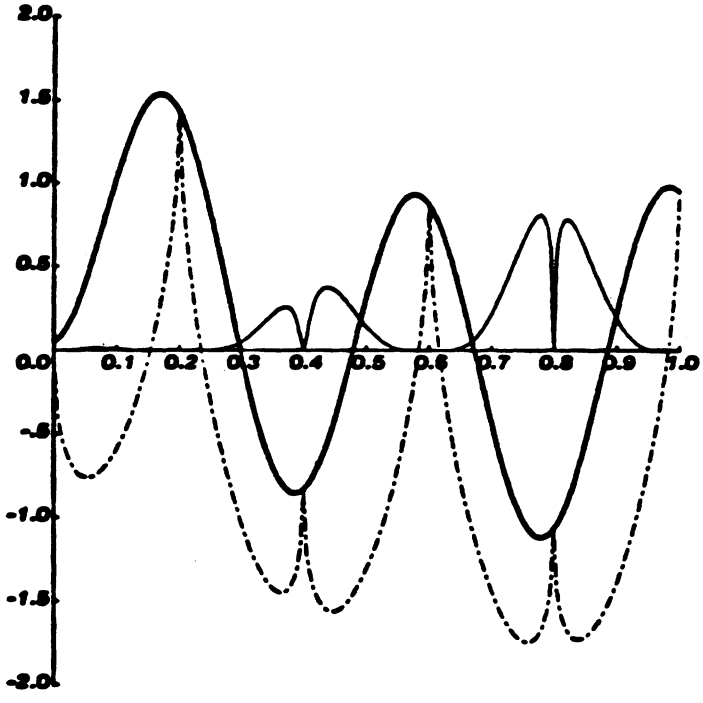

Figure 2-2: Illustration of the objective function and two auxiliary functions f(x) - bold line; $\phi_{k+1}(x)$ - solid; and w_{k+1} - dotdash

$$w_k\hat{}(x) = m_k(x) - c^*_k(x) \qquad a\ [a,b] \tag{12}$$

where $m_k(x)$ and $<r_k(x)$ are defined in (11). Again assume that the observation points have been ordered as in (10). A typical shape for this auxiliary function is shown in Fig. 2-2. Note that for a given $\bar{x}$, the value of $^{w}_{k+1}(*)$ depends on the length of the subinterval to which $\bar{x}$ belongs and on the values of f(x) at the ends of this subintervaL Moreover, for a given constant value c, the (conditional) probability that $f(x) > w_k\hat{}(x)$ is the same at every x *e* [a,b] (eg., for c=2, this probability equals 97.7%). Further, notice that the function $(-w_k\hat{}(x))$ has all the properties listed for $\hat{}_{k<H}(x)$ in Section 2. We propose that instead of searching for a maximum of $\hat{}, 0c)$. we search for a minimum of $w_k\hat{}(x)$, using a procedure similar to the Zilinskas method. Towards this end, we first determine the minimum value of $w_k\hat{}(x)$ in each of the subintervals defined by (10), and then select as the next observation point, $x^{k}*\backslash$the point which corresponds to the least of these minima. The computational effort associated with this procedure is clearly less than that required to solve (9). Furthermore, as will be seen in Section 5, the examples show close agreement between results obtained from the solution of (9) and those obtained using (12).

Unfortunately the use of (12) introduces the risk of missing the global minimum, even when k->oo. This can occur if the minimum of $^{w}_{k+1}(x)$ in some subinterval is greater than $f(x^{opuk})$ became no observation points will then be chosen in this subintervaL While this risk will decrease for larger values of c, larger values of c usually increases the number of obsevations required. A possible solution would be to adjust the value of c at every optimization step.

Note that because of our choice of auxiliary function, the new observation point and the interval within which it is located, represents, in some sense, the best tradeoff between the uncertainty of the interval as characterized by $<?_k(x)$, and the expected value of the optimum as characterized by $m_k(x)$.

4. EXTENSION TO MULTIDIMENSIONAL GLOBAL OPTIMIZATION

Clearly in order for global optimization techniques to be suitable for circuit design, they must be able to solve multidimensional problems. The extension of the Kushner and 7iiinskas methods to higher dimensions requires consideration of multidimensional stochastic processes, along with their complex conditional parameters, thereby making computation intractible. However, the method introduced in Section 3 can be extended relatively easily. As in the one-dimensional case, we will define an auxiliary function, which has the properties of the auxiliary function defined in (12). Further, we assume that we have a set of observation

points which caa be used to divide the feasible region A into a set of nonoverlapping subregions. We will then compute the minimum value of the auxiliary function in every subregion of A and choose as the next observation point the point which corresponds to the least of these minima. This new observation point will then be used to futher subdivide the feasible region and the process is repeated

4.1. DIVISION OF FEASIBLE REGION A INTO SUBREGIONS

Recall that we have already assumed that the feasible region A is a hyperrectangle whose vertices are determined by the lower and upper limits of the values of the designable parameters. While these vertices could be used as the initial set of observation points, we choose not to for two reasons. First, the number of such vertices grows rapidly as the number of dimensions of x increases. Second, the feasible region A can be huge, and trying to subdivide an extremely large region could be computationally expensive. As an alternative, we propose to begin with a set of (n+1) arbitrarily choosen observation points in A, which defines a simplicial approximation to A. We will then employ either a simplicial subdivision scheme, to refine this approximation, or a polyhedral inflation scheme, to allow for improved coverage of A, if necessary. We now give the details of this procedure.

Let x^1, x^2, ..., x^{11*1} be the set of (n+1) randomly chosen observation points in A, and let D^{0*1} denote the polytope (actually at this point a simplex) whose vertices are these points.*** Now consider the observation point, $x^{\circ*2}$. This new observation point will fall either inside the polytope $D"^{*1}$ or inside its complement"5^{11*1}. Note that $A=D^{n*1}$ \J~tr[1] and ET^1 f]^[1] = 0. In the first case, we will refine the approximation to A by further subdividing D^{0*1} into a set of nonoverlapping simplices, S_j^{0*1}, j = 1, 2, ..., n+1 which are defined by x^{n*2} along with all combinations of n vertices of $D"^{*1}$. Note that if x^{n*2} falls on the face of D^{n+1} degenerate simplices may arise which are ignored. This newly subdivided $D"^{*1}$ will be called D^{n+2}. If x^{n*2} falls inside of $\bar{D}^{11*1}$, we will see in Section 4.3 that because of our choice of the auxiliary function x^{D*2} must be a vertex of A. We now inflate the approximation to A by adding to D^{0*1} the simplices defined by this vertex of A and the vertices of the faces of the polytope $D""^1$ as seen by this point This new approximation is denoted by $D^{\circ*2}$ (see [15] and [16] for techniques to achieve this step). These concepts are illustrated in Fig. 4-1.

The above procedure can in general be repeated to generate approximation D^{**1} from approximation D^k given the observation point x^{k+1}. Before leaving this section, it is convenient to introduce some additional notation which will be needed below. Observe that after a

***Note that polytopcs D^1, D^2, ... D^n are undefined.

number of subdivision and/or inflation steps, we have an approximation to A, denoted by D* and its complement "5", such that A = D" U "5" but D" H *D" = 0. Further, D* is composed of I^k nonoverlapping simplices, Le..

$$I^* = \bigcup_{i=1} S_i^k$$

We now wish to consider the decomposition of $\bar{D}^*$ into J^k subregions G_j^k, j=l, 2. . . . J_v^k where J^k is the number of faces of the polytope E^ and G_j^k is defined as follows. Observe that the hyperplane defined by the j^{1h} face of polytope *I?* seperates two half spaces, one of which does not contain &. If we denote this half space by H_j^k, then G_j^k is defined as

$$G_j^k = H_j^k \bar{H} \wedge , \quad j = l\,2 , \; -J_k \tag{13}$$

4.2. DEFINITION AND MINIMIZATION OF THE AUXILIARY FUNCTION

We obtain the auxiliary function of the n-dimensional vector x by generalizing expressions (11) and (12). Towards this end note that the <r(x) component of w(x) in one dimension is an ellipse equation over the subinterval which has its maximum value at the center of the subinterval and zero value at the end points of the subintervaL Observe that the expression (11) may be transformed to the form

$$\sigma^2(x) = \frac{\sigma^2}{2\rho}[\rho^2 - (x - x^a)^2]; \qquad \text{where } {}_P = \frac{x^U - x^L}{2}, \quad x^a = \frac{x^u + x^L}{2}$$

x^u, x^L - upper and lower endpoints of the subintervaL

Extension of (13) to the n-dimensional case yields

$$<r^2(x) = \frac{\sigma^2}{2\rho} \; C/iMx\text{-}x^1)^1 \; (x\text{-}x^1)] \qquad x \ll ACR^n \tag{14}$$

In order for *aHx)* to retain features of $*^2(x)$, we must assume that the simplices which comprise the approximation to A are canonical. A canonical simplex is one in which the distances between every pair of vertices are identical. The center of this canonical simplex, x^a,

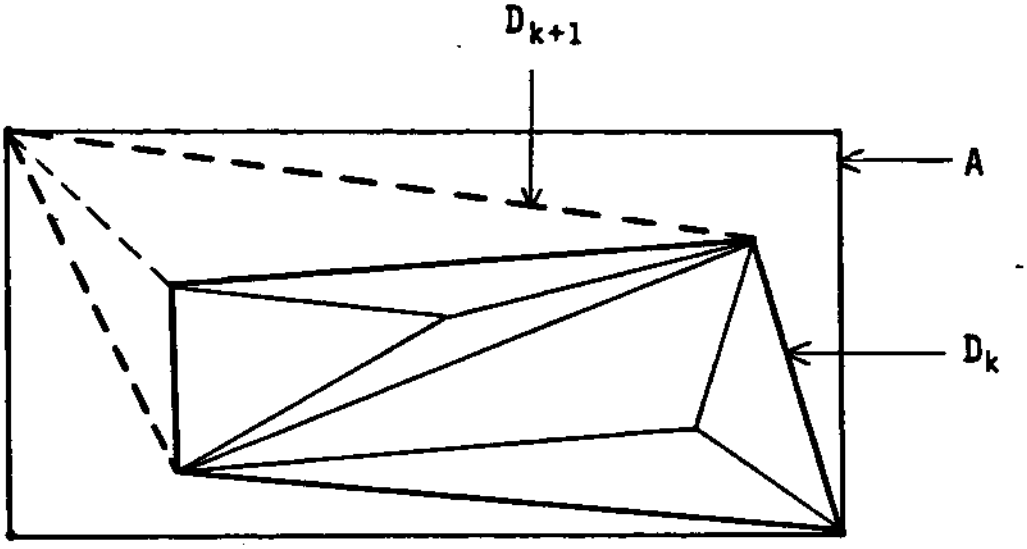

Figure 4-1: Procedure for subdividing and inflation of the approximation to A

is the average of the coordinates of its vertices; *p* is the distance of any vertex from the center. Unfortunately, it is highly unlikely that all simplices which comprise D* will be canonical However, we can transform every simplex into a corresponding canonical simplex of equal volume. This transformation is described in Appendix A. The volume of the simplex is a measure of the uncertainty of f(x) over a given subregion and for this reason it should be unchanged. Expression (14) is attractive since, as will be seen below, the minimum of w(x) may be found analytically (see (17) and (BID).

We can now define the auxiliary function within each of the subregions which make up A as

$$w_i^{k*}(x) = m_i^k(x) - c\,\sigma_i^k(x) \qquad (15)$$

where j denotes either one of the simplices which make up D* or one of the subregions in "b* as defined by (13) and the functions m_i^* and σ_i^* are defined as follows. Consider first that part of the region A contained in D\ which has been further subdivided into the simplices S_i^*. i=l, 2, ... I_k. Let v_n, v_p. . . . $v_{i,n}$^ represent the vertices of the i* simplex; f_{Jf} .... $f_{i.}$^ be the values of the objective function at these vertices; $\tilde{S}_i^k$ be the equivalent canonical simplex associated with S_i^* and $?_{j,}$, $?_{p}$. . . . $?_{p,M}$ the vertices of the canonical simplex. In what follows, we take f_{ij} to be the objective funtion value at the vertex $?_{ij}$ as well as v_{ij} We now define m_iHx) to be specified by the equation which characterizes the n+1 dimensional hyperplane which passes through the points defined by $(f.\ V.J)^7$, $(f...\ 9\_^T)^T$, ... $(f\ ;\ 7.\ ,^T)^T$. In particular, we can express mHx) as:

$$m_i(x) = (m)^T x > \bar{m}_i^k$$

where the n-vector m contains the last n components of the normal to the hyperplane defined above and ^ is the distance of this hyperplane to the origin. Finally, $(a\backslash x))^2$ is a hyperellipsoid which circumscribes the equivalent canonical simplex $\tilde{S}_i^*$.

Now consider that part of the feasible A made up by the subregions G_j^k as defined by (13). Each of these subregions is defined by a face of the polytope D*. Let v_{jl}^*. . . . v_{jm}^* denote the vertices which define the j* face of D\ We choose $m_j$tx) to be a constant outside of D^k such that

$$m_j^k(x) = \min_i [f(v_{ji}^*)]$$

We define $(<r_j!W)^2$ in this region as a linear function of the distance, d, from the face of D \ Le.

$$(\sigma_j^k(x))^2 = <r^2 d$$

Parameters c and a^2 are discussed more fully in Section 4.4.

Note from the above definitions, that outside of the polytope face, Le., within some subregion G_j^k, the function $w_j^{**}(x)$ becomes minimal at one of the vertices of the hyperrectangular box A. This vertex may be easily detected by testing the vector normal to the face of the polytope defining the particular subregion G_j^k, and evaluating w(x) at this point However, minimization of w_i^{**1} (x) over the canonical simplex requires solution of the constrained optimization problem:

$$\text{minimize: } w^{\wedge}(x) = (_i m\ W_i - \frac{c}{2\rho} {}^{\wedge}L_f[/>^2-(x-x^a)^T(x-x^a)]^{1,2} \quad (16)$$

$$\text{subject to: } (^{\wedge}_j Vx-1 \pounds 0 \qquad j=1,2,...,n+1,\ n>1$$

where **$_{ij}$) is a vector normal to the j* face of simplex $\tilde{S}_i^*$ with origin at x*.

Although $w_i^{**'}(x)$ in this case is strictly convex, there is no analytical^solution to problem (16). In most cases however, the solution of (16) falls inside the simplex S_i Therefore, as an alternate to solving (16) we can consider

$$\text{minimize } w_i^{**1}(x) \quad (17)$$

where w_i^{**1} (x) is as defined in (16). The solution of (17) is

$$x^{\circ} = x - \frac{*}{t(cV/2\sqrt{W_{n+1}})^{T} m_j]^{1/2}} \qquad (18)$$

(Note that (18) is a generalized form of the solution for problem (12), defined in Section 3).

Next we check whether or not x° is outside the simplex under consideration by solving for the X in the following problem:

$$\sum_{j=1}^{n+1} \lambda_j v_{ij} = x^{\circ} \qquad (19)$$

and

$$\sum_{j=1}^{n+1} \lambda_j = 1$$

If all X_i, i=l,2^..,n+l are non-negative, then x° is inside the simplex. If x° falls outside the simplex, then we solve the following constrained problem:

minimize: $w^{**1}(x)$

subject to: $g_j(x) = (*J^{T}x - l = 0 \qquad j=l,Z...J \qquad (20)$

where $w_i^{**1}(x)$ is as before and the constraints $g_j(x)$ correspond to the J negative X_j which results from the solution of problem (19). An analytic solution to this problem is given in Appendix E

In higher dimensions, i.e., for n > 2, the solution of (20) may also fall outisde the simplex. Thus problem (19) must be solved using the solution obtained for (20). If some of the X_i are still negative the appropriate constraints are appended, and (20) is resolved

Since x^0 is the solution point for an equivalent canonical simplex, the solution point in the original simplex is found by backward transformation (Appendix A).

4.3. MAIN ALGORITHM

We now summarize the procedure for global optimization. Assume that a set of (n+1) observation points within the feasible region A have been choosen and that an objective function evaluation has been performed at each of these. These points define the simplex S_1^{n+1}, which constitutes the initial approximation, D^{n+1}, to A. furthermore, as new observation points are formed, the approximation to A is refined in such a way so that it remains composed of nonoverlaping simplices. Thus, the k^{th} step in the procedure involves:

1. Minimization of the auxiliary function (15) over each of the simplices, S_j^k, that make up the approximation, D^k (which involves repeated solution of (17), (19) and (20)) and over each of the regions G_j^k in $\overline{D}^k$, as defined by (13) (which involves evaluation of the objective function at a vertex of A).

2. Selection as the new observation point, x^{k+1}, that point which yields the least of the minima found in 1.

3. If the uncertainty of the region which contains the new observation point is acceptably small, then take this region as being the neighborhood of the global minimum, and x^{k+1} as the initial point for a local optimization procedure. If the undertainty of this region is too large, continue to the next step.

4. If x^{k+1} falls within a simplex which makes up D^k, subdivide this simplex using x^{k+1} as a new vertex. If x^{k+1} falls in $\overline{D}^k$, i.e., at a vertex of A, then inflate the polytope D^k to include x^{k+1} as a vertex.

4.4. PARAMETER EVALUATION

We now consider evaluation of the parameters σ^2 and c. As for the one-dimensional case in the expression for w(x), two constant values σ^2 and c exist. In the one dimensional case, σ^2 is a Wiener process parameter and corresponds to differences of observation values. At this point, we generalize the assumption made in the one dimensional case, namely that f(x) is a realization of a certain stochastic process. As shown by Mockus [10] the most suitable stochastic model is a multidimensional Gaussian function defined on A with expected value μ and covariance matrix K whose entries are defined as:

$$K_{x^i x^j} = \sigma^2 \prod_{t=1}^{n} \left(1 - \frac{|x_t^i - x_t^j|}{\Delta_t} \right) \qquad i, j=1,2,\ldots,n \tag{21}$$

where

$$\Delta_t = x_t^U - x_t^L$$

Estimators of parameters /i and a^1 may be obtained using the method of unbiased likelihood estimation [10]. Let the number of observation points equal M. We obtain first the estimator ^ of expected value /i

$$\hat{\mu}^* = \frac{1^T K^{-1} f}{1^T K^{-1} 1} \tag{22}$$

where

f — is the column vector of M function observations

1 — is the unit M-vector

The estimator $\hat{\sigma}^2$ of σ^2 is as follows:

$$\hat{a}^2 = \frac{1}{M-1}(f - 1\hat{\mu})^T K^{-1} (f - 1\hat{\mu}) \tag{23}$$

Note that this method of estimating a^1 needs further investigation since the auxiliary function w(x) was created by an analogy to the w(x) in the one-dimensional case, and not by analogy with the stochastic process. Nevertheless, formula (19) keeps the information about the range of differences of f(x) values and for this reason appears to be useful.

The meaning of the constant c remains unchanged in higher dimensions. Further, the study of the behavior of the proposed algorithm as c varies should bring further improvement in its efficiency.

5. NUMERICAL EXAMPLES

5.1. ONE-DIMENSIONAL CASE

Three different one-dimensional multiextremal functions were considered:

1. $f_1(x) = 2(x - 0.75)^2 \cdot \sin(5\pi(x - 0.4)) - 0.125 \qquad 0 \le x \le 1$

   which has a global minimum at x = 0.7795 with f_{min} = -1.1232287

1 $f_2(x) = \min [f_{2J}(x), tjx)]$ $0 \le x \le 1$

where

$f_{2J}(x) = 2(x-0.75)^2 + \sin(8\pi x - 0.5\pi) - 0.125$

$-1.25 \cdot 75(0.17 - x)$ $x \le 0.17$

$tjx) =$

$-1.25 + 35(x - 0.17)$ $x > 0.17$

which has a global minimum at x=0.17 with f = -1.25 and two local minima at x = 0.75 with f(x) = -1.125 and x = 0.99842 witff"f(x) = -1.0007867

3.

$$f_3(x) = -\sum_{j=1}^{5} [j \sin t(j+1)x+j]J$$ $-10 \le x \le 10$

which has three equal minima at the points -6.77457, -0.49139 and 5.79179 with f = -1103125, and one local minimum at x = 4.5577 with f(x) = -9.4947. """

In all of these examples, six observation points were initially chosen in the region A which resulted in the region A being divided into 5 equal subintervals. Table 5-1 compares the new observation points chosen by the original Bayesian method .(i.e., Silinskas) and the simplified method, for different values of c It was decided a priori to fix the number of observations at 31 Table 5-2 shows the two best results for different values of c and the number of observations required to achieve these results.

5.2. TWO-DIMENSIONAL CASE

Two different two-dimensional multiextremal functions were considered:

1. Branin's function.

^Ocytoty-bx^x-d)^2 + h(l-f)cos(x) + h

where a=l. b=5.1/4»^{2}, c=5/», d=6, h=10. f=l/8»

-5 £ x * 10, 0 £ y £ 15

This function has three equal global minima at points (-3.14159, 11275), (3.14159, 1275), (9.42478, 1475) with f^ = 0.397887.

1 Goldstein's and Price's function [17]

$f_2(x,y) = [1-Kx+y+1)^2(19-14x+3x^2-14y+6xy+3y^2)]$ x

[illegible]	c	»,— M^	[illegible]
•*• [illegible]	1 I 3 T	[illegible] JfTO [illegible] JSF	[illegible]
•M [illegible]	1 t 3 T	JHO Jtti J09§ * * .	j r a M I *mm* *ma*
r [illegible]	1 3 3 T	[illegible] [illegible] ra* TJO*	*%m* JJU [illegible]

Table 5-1: Comparison of observation points determined by the Bayesian and simplified methods

Function	n	Bayesian Method			Simplified Method		
		x	f(x)	obs.	x	f(x)	obs.
	1	.7700	-1.1332	35	.7720	-1.1332	20
		.7700	-1.1332	7	.7700	-1.1332	16
	2	.7702	-1.1332	9	.7704	-1.1332	9
		.7771	-1.1327	28	.7810	-1.1328	28
	3	.7707	-1.1332	30	.7707	-1.1332	21
f-uv		.7700	-1.1331	32	.7700	-1.1331	29
	7	.7010	-1.1327	28	.7012	-1.1328	29
		.7725	-1.1167	28	.7720	-1.1162	29
	1	.8000	-1.0000	0	.8000	-1.0000	16
		.8070	-1.0000	11	.8000	-1.0000	32
	2	.8000	-1.0000	16	.8000	-1.0000	32
		.8070	-1.0007	9	.8000	-1.0000	12
	3	.7984	-1.1330	32	.8000	-1.0000	21
		.7984	-1.1340	35	.8002	-1.0000	32
	7	.7400	-1.1347	17	.7400	-1.1330	16
		.7400	-1.1210	32	.7320	-1.1004	25
	1	5.7000	-11.0040	5	5.7017	-11.0010	24
		5.0030	-0.4300	9	5.7000	-12.0011	10
	2	-.4000	-12.0000	10	5.0401	-11.0703	32
		-.4010	-12.0171	16	5.0430	-11.0000	31
	3	-.4005	-12.0007	14	5.0330	-11.0000	32
		-.4040	-12.0030	11	5.0005	-11.0000	21
	7	4.0007	-0.4041	35	4.0030	-0.4000	21
		4.0000	-0.3400	30	4.0004	-0.4000	33

Table 5-2: The results of the optimization by the Bayesian and simplified methods

$$[3(W2x-3y)^2(18-32x+12x^2+48y-36xy+27y^2)]$$

$-2 \pounds x \pounds 2, \qquad -2 * y \pounds 2$

This function has a global minimum at the point (0.000, -L000) with $f_{Bia} = 3$ and the local minima as given in the following uble:

x	-.600	-.398	1.800	L200
y	-.400	-.602	.200	.800
f(x.y)	30	35	84	840

The algorithm presented in Section 4.4 was used to find the neighborhood of the global optimum. The two best results, after 36 observation points were found for different values of c, are shown in Table 3.

5.3. A FIVE-DIMENSIONAL EXAMPLE

To illustrate the performance of the proposed technique for higher dimensional problems, consider the circuit of a constant voltage reference described by Heydemann, et al [18] shown in Fig. 5-1. The circuit is designed to keep the output voltage v^ constant in spite of variations in the set of technological parameters such as transconductance, threshold voltage, channel width absolute tolerance, channel length absolute tolerance and temperature.

The designable parameters are the channel length of each of the MOS devices, denoted by $(x_{J9}...,x_s)$. The following five-dimensional optimization problem can be formulated:

minimize $f(x_r...x_5) = (V_{out} - 5)^2$

subject to

$$\begin{aligned} 15 &\le x_1 \le 100 \\ 15 &\le x_3 \le 150 \\ 15 &\le x_2 \le 150 \\ 15 &\le x_4 \le 200 \\ 15 &\le x_5 \le 200 \end{aligned} \qquad \text{(length in microns)}$$

Heydemann solved this problem using a grid technique [18] and found, after 2500 circuit analyses, 4 different global minima at which f(x) = 0.0.

Since the method proposed in this paper determines only a neighborhood of the global optimum, it is hard to compare the efficiency of both approaches. However, after only 100 circuit analyses, we obtained a reasonable neighborhood of the global optimum. Any point in

Function	#	x	y	f(x,y)	obs.
$f_1(x,y)$	1	Milt	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]
	2	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]
	3	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]
	7	[illegible]	[illegible]	[illegible]	ir
		[illegible]	[illegible]	[illegible]	*
$f_2(x,y)$	1	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]
	2	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	ntwr	[illegible]
	3	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]
	7	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]

Table 5-3: Results for the optimization of functions of two variables

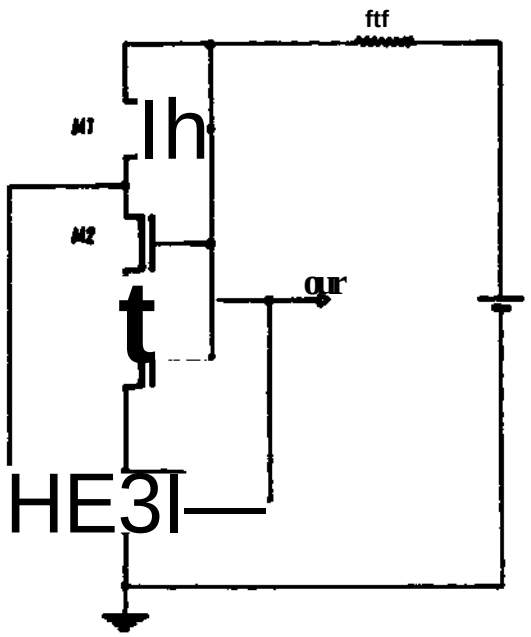

Figure 5*1: Circuit example.

this neighborhood could be used as the initial point for any local optimization technique. Table 5-4 summarizes the results of applying the proposed method for global optimization and five best estimates of the optimum for various values of c.

6. CONCLUSIONS

A new method for seeking the global optimum in circuit optimization problems has been developed which requires a relatively small number of function evaluations and has a low computational overhead. Unfortunately, we cannot establish a strict relationship between the dimensionality of the problem, Le. the number of design parameters, and the number of function evaluations required for its solution. However, based upon a number of examples, it appears that if the number of function evalutions is about 20 times the number of design parameters, a reasonable neighborhood of the global optimum can be found. Increasing the number of function evaluations 2 or 3 times does not *seem* to significantly change the results. This neighborhood of the solution is a good point at which to employ a local optimization method The efficiency of proposed algorithms depends strongly on the techniques used to divide the parameter space. Hence, further investigation into techniques for this purpose is suggested To decrease the risk of missing a global optimum further study of how best to choose the constant c is needed

c	PARAMETERS					$f(x_1,\ldots,x_5)$	v_{out}	# Obs.
	x_1	x_2	x_3	x_4	x_5			
1	25.66	33.69	124.6	158.3	150.7	0.000029	4.995	41
	35	21.22	116.8	149.5	123.7	0.001338	4.963	69
	27.66	20.9	94.16	122.5	142.6	0.001619	5.04	79
	27.75	20.87	94.22	172.5	92.68	0.003964	4.937	78
	27.35	20.59	128.1	125.2	96.28	0.00405	5.064	82
2	28.53	19.19	95.74	134.3	103.6	0.001256	4.965	100
	25.28	36.89	124.8	161.2	154	0.002974	4.945	47
	23.27	15	51.64	168.8	62.76	0.04688	4.783	59
	25.15	30.21	101.6	130.2	99.7	0.06899	4.737	41
	40.66	36.78	100.4	160.9	187.7	0.0818	4.714	67
3	25.16	37.96	124.9	162.2	155.1	0.004965	4.93	49
	40.13	37.88	101.1	162	187.7	0.086	4.707	78
	26.14	15	48.47	166.9	59.23	0.1303	4.639	63
	25.18	33.21	100.3	127.8	95.23	0.1416	4.624	44
	15	15	150	200	15	0.1429	5.378	18

Table 5-4: Results for the optmization of the circuit example

APPENDIX A

A.1 SIMPLEX TRANSFORMATION

Consider a simplex defined by the vertices p^1, *j?,...,?*[1]*. Assume that the origin of the parameter space is at p"*[1] and define the matrix P as

$$P = \left[p^1 - p^{n+1} \quad p^2 - p^{n+1} \quad \cdots \quad p^n - p^{n+1} \right] \tag{Al}$$

Let the matrix

$$S_o \bullet \left[< \quad s_o^2 \quad \cdots \quad s_o^n \right] \tag{A2}$$

characterize an a priori selected canonical simplex with the origin at s_o^{*+1}. Observe that for any scalar k > 0, a matrix

$$S = kS_o \tag{A3}$$

also corresponds to a certain canonical simplex.

We wish to transform the simplex P into the canonical simplex S whose volume is the same as P, so that

$$\mathbf{detS = detP} \tag{A4}$$

Let the square matrix T correspond to the transformation of P into the known simplex S_o

$$TP = S \qquad \text{(A5)}$$

From (A3). (A4), and (A5) we have

$$S = (\det(T^{-1}))^{-1/n} S_{\theta} \qquad \text{(A6)}$$

If S* is a certain point in the canonical simplex S then the corresponding point p* in the original simplex is defined by the backward transformation

$$p^* = p''^{*1} + (\det CT^{1}))^{-1/n} r's^* \qquad \text{(A7)}$$

APPENDIX B

B.1 CONSTRAINT OPTIMIZATION OVER A SIMPLEX

Let S denote a certain canonical simplex defined by (Bl)

$$(f_{\cdot})^T x \ S \ 1 \qquad i=1X.jJ+l \tag{Bl}$$

where $j._1$ are the vectors normal to the simplex faces with origins at its center x*. Assume that the solution of an unconstrained optimization problem over S (see (16)) falls outside j faces. In this case, solution x° of the following constrainted optimization problem is required:

$$\text{minimize} \quad w(x) = m^T x + m_o - kOl^2 - (x-x^R)^T(x-x^R))^{1/2}$$

$$\text{subject to} \quad g_j(x) = (*_j)^T x - l = 0; \qquad j=1.£\sim.J \tag{B2}$$

Since w(x) and $g_j(x)$, j=1.2....J are convex, the Kuhn-Tucker necessary and sufficient optimality conditions for (B2) are:

$$\nabla w(x^o) + \sum_{j=1}^{J} u_j \nabla g_j(x^o) = 0$$

$$u_j g_j(x^o) = 0. \qquad j=1.2....J \tag{B3}$$

Hence we get:

$$m + \frac{kx}{CR' - C V l' h V .]^{\wedge} \ \wedge \ ",'ii \ " \bullet} \qquad \sum_{} \tag{B4}$$

Multiplying (B4) by $(»3_j^T$, j = 1, 2,.... J, we gee

$$(\eta_j)^T m + \frac{k}{[R^2-(x-x^R)^T(x-x^R)]^{1/2}} + (\eta_j)^T \sum_{j=1}^{J} u_j \eta_j = 0 \qquad j=1.2....j \tag{B5}$$

After some further manipulation we obtain:

$$H \begin{bmatrix} u_2 \\ \vdots \\ u_j \end{bmatrix} = -(H,m + u_1 H_d \eta^1) \qquad \text{(B6)}$$

where

$$H_d = \begin{bmatrix} (\eta^1)^T - (\eta^2)^T \\ (\eta^1)^T - (\eta^3)^T \\ \vdots \\ (\eta^1)^T - (\eta^j)^T \end{bmatrix}$$

$$H_r = \begin{bmatrix} \eta^2 & \eta^3 & \cdots & A \end{bmatrix}$$

and $H = H_d H_r$

Hence

$$\begin{bmatrix} u_2 \\ \vdots \\ u_j \end{bmatrix} = -(Km + u_1 K \eta^1)$$

where $K = H^{-1} H_d$ (B7)

From (B4) and B7) we have

$$m \cdot \frac{kx}{[R^2 - (x - x^R)^T (x - x^R)]^{1/2}} * ny - H_r(Km + UjK,^1) = 0 \qquad \text{(B8)}$$

Defining the vectors % and 9' as

$$\tilde{\mathbf{m}} = (I - H_r K)\mathbf{m}$$

and

$$T = (I - H_r K)^{-1} \tag{B9}$$

and multiplying 038) by $(*')^T$ we obtain

$$\frac{k}{[RMx - x^R)^T (x - x^R)]^{1/2}} \cdot J^T (\& + u T) = 0 \tag{B10}$$

and the optimal point x°

$$x^\circ = x^R - \frac{R(\wedge \cdot u, \wedge')}{[k^2 + (\tilde{m} + u_1 \tilde{\gamma}^1)^T (\tilde{m}) + u_1 \tilde{\gamma}^1)]^{1/2}} \tag{B11}$$

The multiplier u_t is obtained from solution of the quadratic equation:

$$(\gamma^1)^t R(\tilde{m} + u \wedge^1) \cdot (ir^* * (ft * upT)^T (\& \wedge u \wedge^1))^{172} = 0 \tag{B12}$$

The distance between the solution x° from x^R is less than a radius R, although, for (n>2) it may be outside the simplex S again. The procedure described in Section 4.3 is then recommended. It should be noted that if J=l then & = m and $7^1 = T$ -

REFERENCES

[1] Lightner, M.R. and Director, S.W.
Multiple Criterion Optimization for the Design of Electronic Circuits.
IEEE Transactions on Circuits and Systems CAS-28(3):169-179, March, 1981.

[2] Brayton, R.K. and Director, S.W.
Computation of Delay Time Sensitivities for Use in Time Domain Optimization.
IEEE Transactions on Circuits and Systems CAS-72(12):910-920, December, 1975.

[3] Hachtel, Gary, Brayton, R.K. and gustavson, F.G.
The Sparse Tableau Approach to Network Analysis and Design.
IEEE Transactions on Circuit Theory CT-18(1): 101-113. January, 1971.

[4] Director, S.W. and Rohrer, R.A.
Automated Network Design—The Frequency Domain Case.
IEEE Transactions on Circuit Theory CT-16:330-337, August, 1969.

[5] Dixon, LCW. & Szego, G.P.
Towards Global Optimization.
North Holland Publishing Company, 1978.

[6] Kushner, HJ.
A Versatile Stochastic Model of a Function of Unknown and Time Varying Form.
Journal of Mathematical Analysis and Applications 5:150*167, 1961

[7] Kushner, HJ.
A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise.
Journal of Basic Engineering :97-106, March, 1964.

[8] Zilinskas. A.G.
The One-Step Bayesian method for Seeking the Extremum of the Function of One Variable.
Cybernetics (1):139-144, 1975.
(in Russian).

[9] Torn, A.
A Search-Clustering Approach to Global Optimization.
In *Towards Global Optimization* 2, pages 49-62. North Holland Publishing Company, 1978-

[10] Mockus, LE, Tiesis, V. and Zilinskas, A.G.
The Application of Bayesian Methods for Seeking the Extremum.
In *Towards Global Optimization* 2, pages 118-129. North Holland Publishing Company, 1978.

[11] Archetti, F.
A Probabilistic Algorithm for Global Optimization Problems with a Dimensionality Reduction Technique.
In *Lecture Notes in Control and Information Science,* . Springer Verlag, 1980.
Part 1

[123 Mockus, I.B.
On Bayesian Methods for Seeking the Ekstremum.
Automatics and Computers (3):53-62, 1972.
(in Russian).

[13] Mockus, I.E
On Bayesian Methods for Seeking the Extremum and Their Application.
In *Information Processing,* pages 195-200. North Holland Publishing Company, 1977.

[14] Zilinskas, A.G.
A Method of One-Dimensional Multiextremal Minimization.
Technical Cybernetics (4):71-75, 1976.
(in Russian).

[15] Director, Stephen W., and Hachtel, Gary D.
The Simplicial Approximation Approach to Design Centering.
IEEE Transactions on Circuits and Systems CAS-24(7):363-372. July, 1977.

[16] Director, Stephen W., Hachtel, Gary D., and Vidigal, Luis M.
Computationally Efficient Yield Estimation Procedures Based on Simplicial Approximation.
IEEE Transactions on Circuits and Systems CAS-25(3):121-130. March, 1978.

[17] Goldstein, A.A. and Price, Y.F.
On Descent from Local Minima.
Mathematics of Computation 25(115):569-574, July, 1971.

[18] Heydemann, M., Durbin, F. and Montaron, J.
A Tentative Approach Towards Global Optimization in the DC Case.
In *Proc. of the International Symposium on Circuits and Systems,* pages 847-850.
IEEE, April, 1981.